Future Passenger Air Traffic Modelling: Trend Analysis of the Global Passenger Air Travel Demand Network

Robin Ghosh[1] and Ivan Terekhov[2]
German Aerospace Center (DLR), Institute of Air Transportation Systems, Blohmstraße 18, Hamburg, 21079, Germany

This paper is about the architecture for modelling future global air traffic on city pair level. Here, the focus lies on the initial trend analysis of global demand data required as a foundation for predicting the global demand network based on socio-economic scenarios. We present a 4-layer philosophy to model the future passenger air transportation system in a generic way on a global scale over time until 2050. The evolutions of demand passengers, demand connections, and airfares are analyzed as global aggregates and as distributions over the great circle distance between city pairs. Further, global trends of key network metrics are displayed.

Nomenclature

a	=	Airfare Slope
ACM	=	Aircraft Movements
AF	=	Airfare
ATS	=	Air Transportation System
DF	=	Directness Factor
DPK	=	Demand Passenger Kilometer
F-data	=	Final Data
GCD	=	Great Circle Distance
GDP	=	Gross Domestic Product
OD	=	Origin-Destination based on city pairs
OP	=	Oil Price
P-data	=	Preliminary Data
RPK	=	Revenue Passenger Kilometer

I. Introduction

IN the DLR project "WeCare" climate mitigating effects of operational and technological changes of flight are investigated in the context of the future air transportation system (ATS) on a global scale with a time horizon until 2050. Therefore, at first, a generic model forecasting future air traffic, on network and fleet basis, is required. This will be implemented in a modular environment, called AIRCAST (air travel forecast).

II. 4-layer philosophy

A generic approach as depicted in Figure 1 is necessary to assess a multitude of possible changes from the introduction of a single technology to growth of air travel demand. Pure passenger aircraft fleet models used to assess the global climate impact as in [1] have no spatial quality, but especially the impact of non-CO2 climate agents are highly dependent on the location of emission. This is one example why it is necessary to project spatial network models into the future in combination with a fleet renewal model. Modelling the spatial distribution of global flights over time is relevant to assess the climate impact of the ATS adequately. Such an approach supports policy planning in terms of aviation mitigation strategies and revolutionary new concepts because the climate impact

[1] Research Engineer, Department Systems Analysis Air Transport.
[2] PhD Candidate, Department Systems Analysis Air Transport.

of aviation depends on species, altitude and latitude of emission. [2] Since the 4-layer philosophy is generic it may also be used to consult national governments or city governments, airlines and airports about estimated future demand and aircraft movements depending on the assumed socio-economic scenario. Furthermore, the generic approach allows evaluating the impacts of operational changes like flying slower or the introduction of new technologies up to the launch of a new aircraft programs on stakeholders forming the core ATS. [4]

We introduce the 4-layer philosophy for a generic build-up of the passenger air traffic system of the future. The four layers (see Figure 1) consist of (1) the origin-destination (OD) demand network, (2) the routes network, (3) the aircraft movements (ACM) network, and (4) the trajectories network. Each lower layer builds on the information of the above layers. By demand here, we mean actual realized demand, since this is the only dataset available on global basis. Thus, realized demand is passengers who wanted to fly and actually flew in a given year from one origin to a destination because the conditions were right for them. Detailed definitions of demand in this sense are given in [5]. Our model environment predicts undirected networks on city level (in contrast to airport level). Starting with the theoretically ideal demand network gradually more information concerning the “operational reality” of aircraft deployment is included, e.g. the effects of hub structures, categories of aircraft used on segments, and air space inefficiencies. With decomposition, there is an increasing relevance for technology evaluation and derivation of requirements for new concepts. On the other hand, there is an increasing relevance for scenario development when results and assumptions are more aggregated. The main goal of this philosophy is to link the two major socio-economic drivers GDP (Gross Domestic Product) and population to the future passenger air traffic network in a quantitative way on city pair level. This will allow assessing the growth of the ATS against efficiency and technology improvements dynamically until 2050 on a network and fleet basis. The theoretical background of quantitative scenarios in the context of systems analysis concerning the ATS can be found in [4].

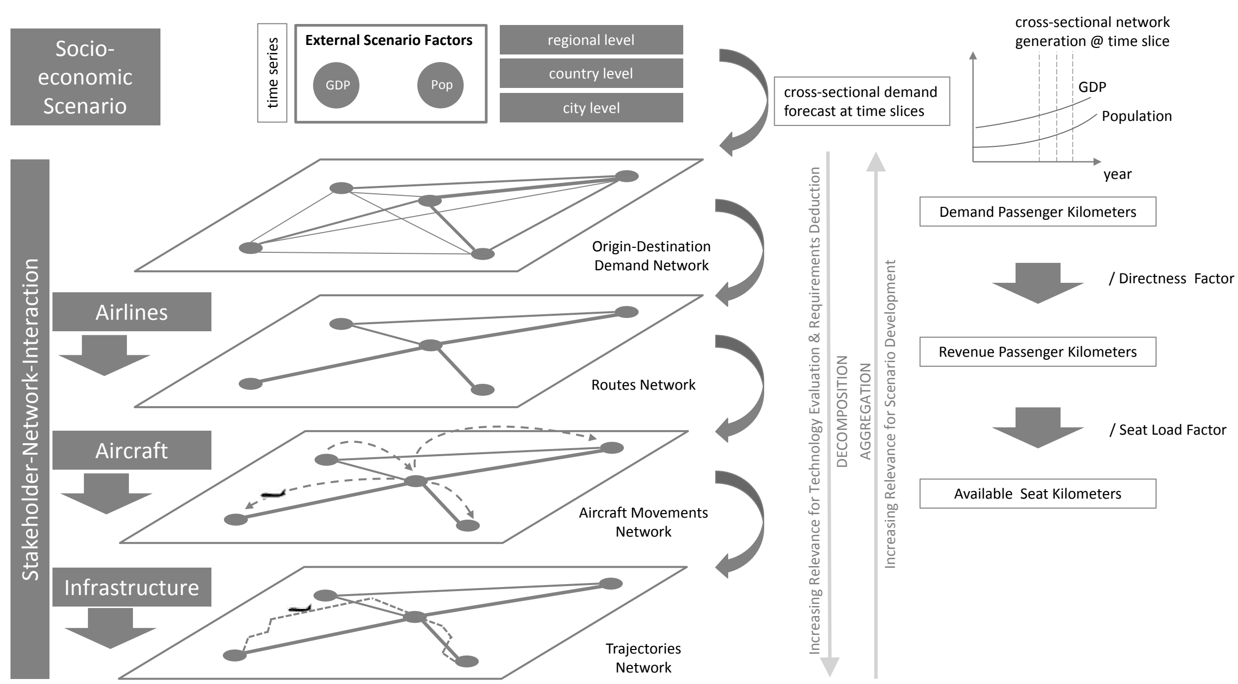

Figure 1: 4-layer philosophy for a generic build-up of the future passenger air traffic system in layers consisting of the origin-destination (OD) demand network, the routes network, the aircraft movements network, and the trajectories network

The combination of dynamic network and fleet information is considered to be highly valuable for systems analysis, technology assessment, policy planning, and master planning of aviation infrastructure. The scientific results of the AIRCAST environment are especially interesting for:

- market forecasts of manufacturer
- strategic planning of airlines (insight in the dynamics of realized demand on city pair level)
- airport capacity evaluation

- aviation-related policy making of city governments, national governments, international organizations (UN, ICAO, IATA)
- dynamic global climate impact assessment of aviation, especially the impact of non-CO2 climate change agents

While the socio-economic scenarios are available as a yearly time series, networks are generated at cross-sectional time slices every 5 years. The main external scenario used at the moment for AIRCAST is Jorgen Randers "2052" forecast [3]. A method and a tool named CITYCAST were developed to decompose the regional Randers forecast to country and finally city level. In total, 4435 cities are included in the forecast model which are 89% of the total amount of cities available in the "Airport Data Intelligence" (ADI) dataset of our base year 2012. 11% of the available ADI-cities from 2012 could not been included because of unavailability of city population data or country GDP data. The subsequent demand prediction model, named D-CAST as described in [5] requires scenario information of socio-economic factors on city level to predict the realized demand worldwide from origin to destination on city level. In the following layers, this aggregation to city level is conserved. Thus, the routes network, defining how many passengers travel on a given segment, and the aircraft movements network, defining how many aircraft of what size category are deployed, will be forecasted on city level. Furthermore, this approach allows ignoring airport choice models at the first step and attaching these models later to AIRCAST. A sample demonstration of the routes to aircraft movements algorithm was published in [6] with a potential use case to forecast aircraft movements at capacity-constrained airports described in [7]. The trajectories network comprises additional information about the amount and location of aircraft emissions as well as time information.

III. Processing of demand data and rearrangement data to city level

Basis for this analysis are realized demand data sets from "Airport Data Intelligence" (ADI) by Sabre Airline Solutions from the years 2002 until 2013. It includes both traffic-types scheduled and charter. Two qualities of data sets are available, preliminary (P-data) and final (F-data). When the study was conducted, preliminary data could be retrieved from 2002-2013 and final data from 2009-2012. We chose the base year for our forecast environment to be the latest available final data at that moment, thus the year 2012. While final data is corrected by mistakes and more accurate, the advantage of the preliminary data set is that it is available for a longer time interval. This is essential for the significance of the trend analysis. The overlapping time interval of the two data sets has been used to analyze systematic deviations and to introduce correction factors for preliminary data on an aggregate level.

Our assumption is that demand exists between two given cities and not between two airports. We aggregated the raw data to city level due to the uncertainty of the future number of airports in a given city and the passenger distribution between them. That means that if more than one airport is located in a city, we merge the passenger flows to city level. Consequently, after the rearrangement process, there is only one demand connection between city A and B, even if both cities have more than one airport and multiple connections in the initial ADI-data. The ADI-data available provides city names according to each IATA code, but does not have unique city names in a given country. There are for example, several cities called "Columbus" in the USA. For a unique identification of a city and for demand forecasting purposes, ADI-data is combined with geographical coordinates. The result of the rearrangement process is the historical passenger demand networks from 2002 until 2013 (example in Figure 17). For the demand forecasting methodology on city level, additionally to coordinates, population and GDP data mapped to cities is required. Therefore, the output of the rearrangement process is combined with global city population and country GDP data, mostly from UN databases. GDP data is than broken down to city level. Thus, all information required for the demand forecasting methodology is available on city level. We erased demand connections with missing necessary socio-economic information because it is not possible to use these connections in the demand forecasting methodology. The overall losses relative to the ADI raw data because of data errors, unavailability of coordinates for IATA codes, merging from airport to city level, unavailability of city population data, and unavailability country population data are below 0,4%, which are about 10,1 million passengers. The merging process from airport to city level reduces the global amount of connections by about 30% compared to the raw data on airport level.

IV. Trend analysis of passengers

An important aspect of building time series of global passenger data is the compatibility assessment of final and preliminary data sets while combining them to analyze a longer period. The calculated passenger numbers on distance intervals are compared to each other for the overlapping period 2009 to 2012 to detect systematic deviations. P-data has systematically less passengers on intervals between zero and 5000 km. The P-year 2011 was identified as an outlier of the time series. Due to this fact, it is not included in the calculation of the correction factor. The average deviation of the P-data of the years 2009, 2010, and 2012 on intervals between 0-5000 km is 5,1% (F-data has 5,1% more passengers). On other intervals of the P-data there is no systematic deviation of the amount of passengers. Therefore, intervals above 5000 km are not corrected. In the base year 2012, more than 90% of all demand connections have a great circle distance below 5000 km. Only 1% of air passenger demand was located on OD city pairs with distances above 12500 km.

A. Aggregate trend

In Figure 2, the amount of global passengers per year is depicted. ADI preliminary and final data are compared to each other and to a third source, the world bank data [8].

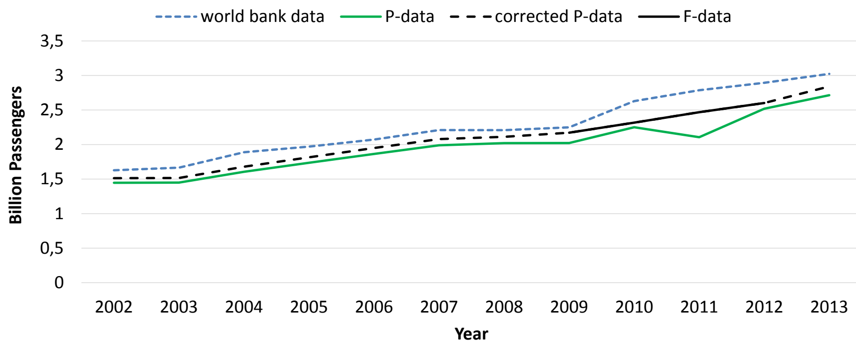

Figure 2: Global air passengers: comparison of preliminary and final ADI-data and world bank data

B. Trend of distribution

In Figure 3 combined final and corrected preliminary data is illustrated as a time series of distributions. The distribution was calculated based on an interval width of 100 km. The graph may be read as follows: on a great circle distance interval between origin and destination from 1000 to 1100 km, a certain volume of passengers traveled in a given year, e.g. 95 million in 2012. The chronological evolution of the passenger demand distribution did show a systematic change during the last ten years. The basic profile of the passenger demand distributions of the various years seems to be stable over time. Nevertheless, the rate of growth varied throughout distance intervals.

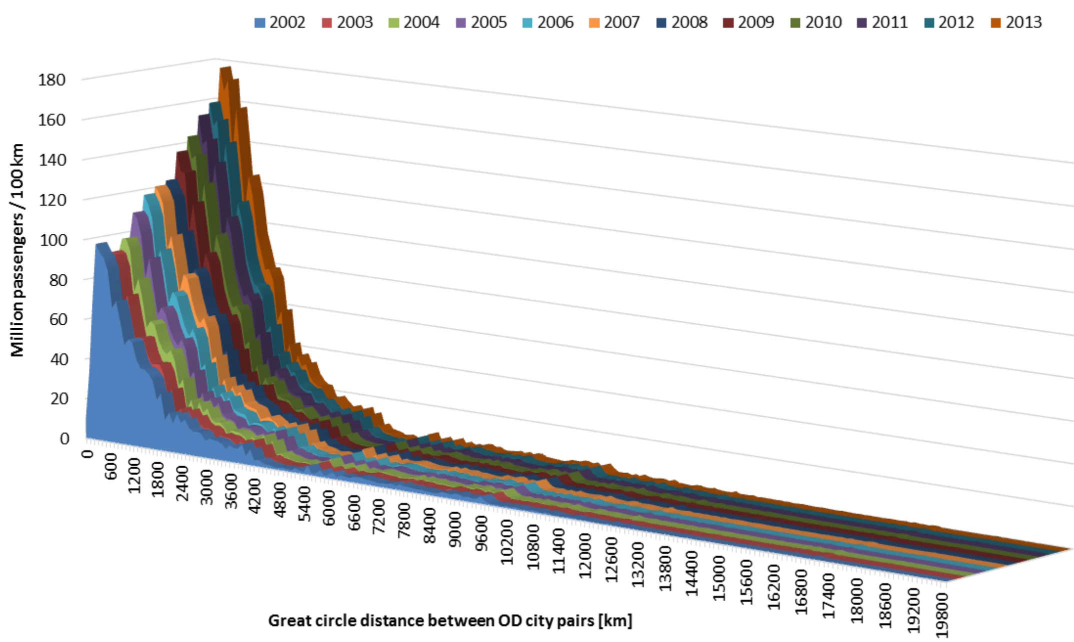

Figure 3: Trend of passenger demand distributions from 2002 until 2013 in million passengers on an interval with 100 km width

C. Interval-specific passenger growth and predicting passenger demand distributions

In order to analyze the behavior of the distribution over time, knowledge about the interval-specific growth is needed (see Figure 4). Comparing the distributions 2002 and 2013, the 11-years trend of interval-specific growth is calculated and compared to global aggregate growth of passengers. In this way, it can be seen on which intervals growth has been disproportionately higher or lower than the global growth. Passenger growth on origin OD distances below 500 km is considerably below global growth. This might be an effect of intermodal concurrence to high-speed rail or improved rail connections in general. In the distance interval between 1500 and 2000 km a first local growth maximum is located. On this interval a considerable amount of flights within Asia could be located, especially within China. There, strong economic growth is fostering the regional air travel demand. On the interval 4500-5000 km, where the second growth maximum is situated, many connections from Europe and Asia to Middle East can be found. Also on these connections, demand grew disproportionately. Other local maxima and minima beyond distances of 6000 km are not discussed here in detail because of the relatively small passenger volumes on these intervals.

The main driver for this kind of analysis is the research question if the passenger demand distribution over distance is evenly scaling up or if the profile might change in the future. As described in the 4-layer philosophy of AIRCAST, an OD demand network needs to be forecasted on city basis. For comparison and consistency checking, we developed an additional approach to predict future passenger demand distributions based on historical interval-specific growth. Since the interval-specific growth varies considerably on different distances, we decided to consider every interval separately to forecast a passenger demand distribution. The method estimates passenger demand on a specific interval as a function of an assumed aggregate passenger growth. This is done for every interval to receive the future distribution, which will differ from a simply scaled-up distribution. Eventually, these distributions need to be compared to the sophisticated approach presented in [5].

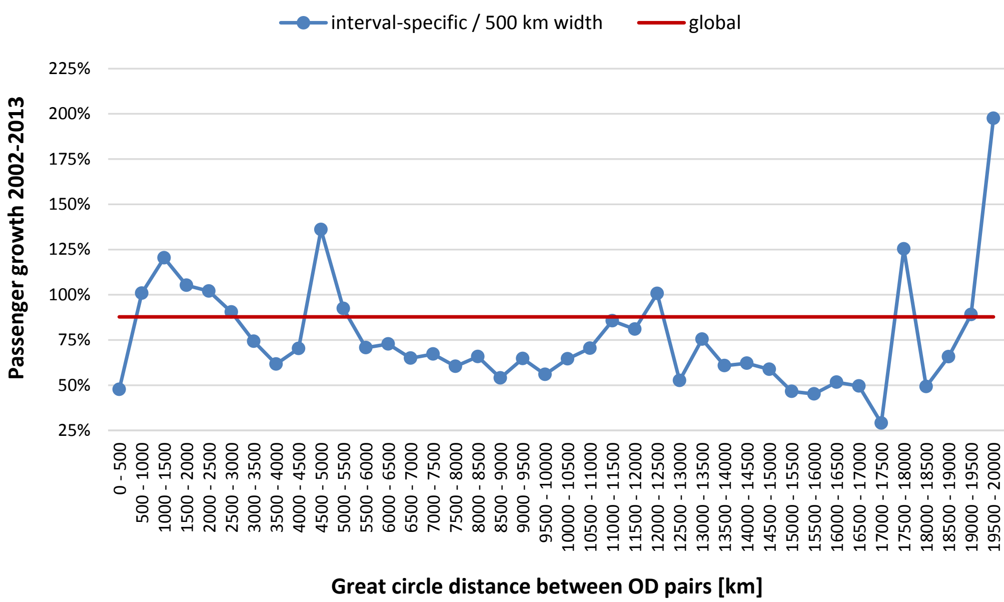

Figure 4: Interval-specific passenger growth 2002-2013 compared to global aggregate growth

In a sample demonstration, the global aggregate passenger growth is assumed to be 5% per year until 2040. For each year of the data set 2002-2013 interval-specific and global aggregate passenger growth is calculated and fitted with a linear regression. Using this method, on an interval between 1000 and 1500 km, for example, a higher yearly growth rate of 6,41% instead of the global 5% is predicted. The growth behavior is consistent with the values over the period 2002-2013 as depicted in Figure 4. Based on this method, a distribution for a given forecast year is estimated. Figure 5 shows the distribution if interval-specific differences are considered compared to an evenly scaled-up distribution. The interval width can be chosen from 100 to 5000 km within the passenger distribution prediction tool, but 200 km width are expected to give most informative results.

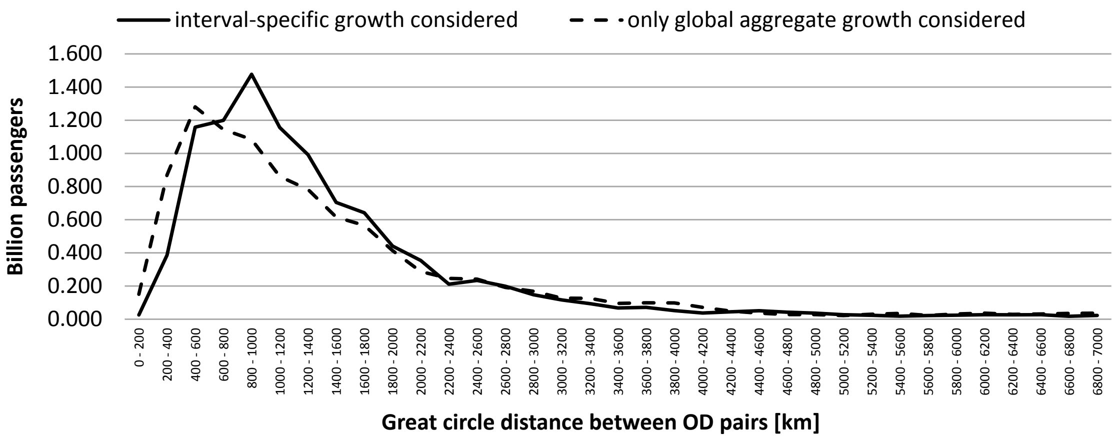

Figure 5: Exemplary demonstration of prognosis method of the passenger demand distribution for the year 2040, GCD between OD pairs 0-7000 km, interval width 200 km

Comparing the two distributions in Figure 5, especially on OD distances below 2000 km, a distinct shift is identifiable. We expect a relative shift from shorter distances (200-600 km) to distance intervals between 400 and 1400 km. Note that in absolute terms all intervals are growing, but some are growing faster than others. The maximum of the distribution is likely to move towards longer distances between OD city pairs. On distances above 2000 km, the differences are not that remarkable because of lower passenger volumes.

As described in Section II, the overall goal is to model the future global air traffic on city pair level. Starting from socio-economic scenarios, future passenger demand networks will be calculated on city level. From these generic networks, passenger distributions can be calculated and compared with the distributions obtained by the method described here considering interval-specific historical growth. The comparison yields information about plausibility and potential model errors or confirms the soundness of the demand network estimation.

V. Trend analysis of connections

D. Aggregate trend

The aggregate trend of the number of demand connections and their distribution is an important information for modelling the future ATS since it strongly influences the structure of the future ATS. Therefore, the demand forecast algorithm would predict new connections beyond the initial data set of the current base year 2012. That means that topology changes over time are considered in AIRCAST. For verification and validation, the historical trend and a projection need to be compared to the connections predicted by topology prediction module of the passenger demand forecast model (D-CAST) [5]. In 10 years from 2002 to 2012, the number of connections in the demand database grew by 19%. In other words, there have been an additional 8700 demand connections every year.

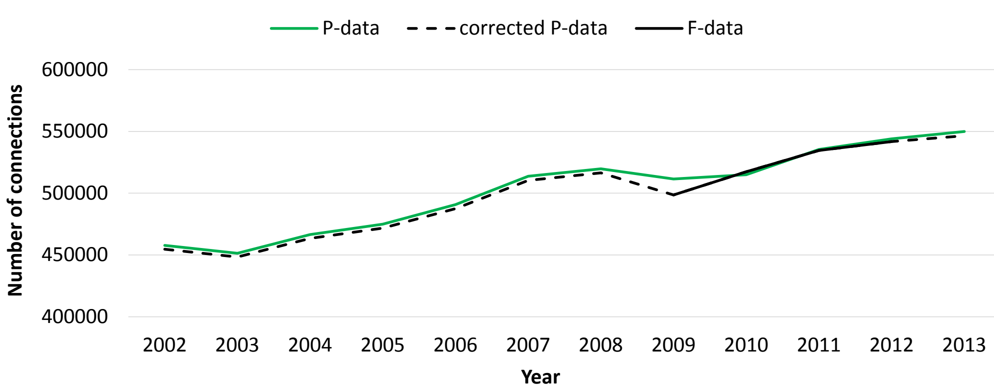

Figure 6: Time series of the total number of OD demand connections. Corrected P-data in green and F-data in blue

The P-data set had in comparison to the F-data set on GCD between city pairs below 5000 km systematically more connections. In order to combine the P-data with the F-data set in one single time series, a correction factor of 0,986 was introduced for P-data on the distance band from zero to 5000 km.

E. Trend of distribution

The corrected realized demand connection distributions of the P-data set 2002-2008 and 2013 are combined with the connection distributions of the F-data 2009-2012. In this way, we get a mixed time series of connection distributions as depicted in Figure 7.

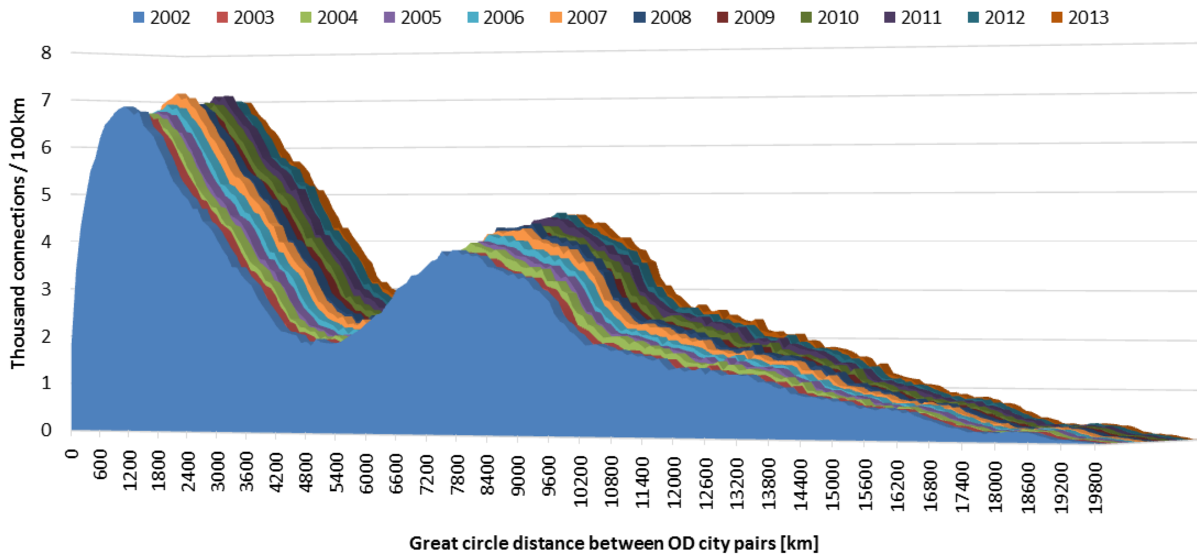

Figure 7: Time series of global connection distributions from 2002 to 2013, interval width of 100 km

The shape of the global connection distributions with two maxima reflects the geographical location of the continents and sea areas concerning origin and destination passenger flows. A local minimum is located at about 4800 km which is approximately the transatlantic distance from Europe to the upper United States east coast. Beyond that distance, more OD demand pairs especially between Europe and the United States can be found. The number of connections on higher GCD rises again to a second maximum at about 8000 km which is approximately the distance between London and Beijing. The 10-years trend shows no major alteration of the general profile of the distribution. We assume that profile will keep its overall shape in the decades to come. Therefore, we will compare this shape to the distributions calculated from forecasted dynamic demand network topologies for plausibility checking. Nevertheless, we analyze interval-specific differences.

F. Interval-specific growth as a function of distance

Comparing the connection distributions of the years 2002 and 2013, we analyze the change of the amount of connections on different intervals of the GCD between OD pairs. The growth is calculated for each interval and opposed to the aggregate growth of connections (see Figure 8).

The growth of connections varies throughout the distance intervals. The number of connections on distances below 3000 km grew less than average. Below 1100 km, the number of connections has even been decreasing. A local maximum of growth on the interval 4500-5000 km correlates with a local maximum of passenger growth on that interval. Here, beside the growth of passengers on existing demand city pairs, new city pair connections arose in addition. This could be, analogously to the passenger growth on that interval, ascribed to new connections from Asia and Europe to Middle East. A second maximum of demand connection growth is located at about 12000 km, approximately the GCD between Shanghai and New York City.

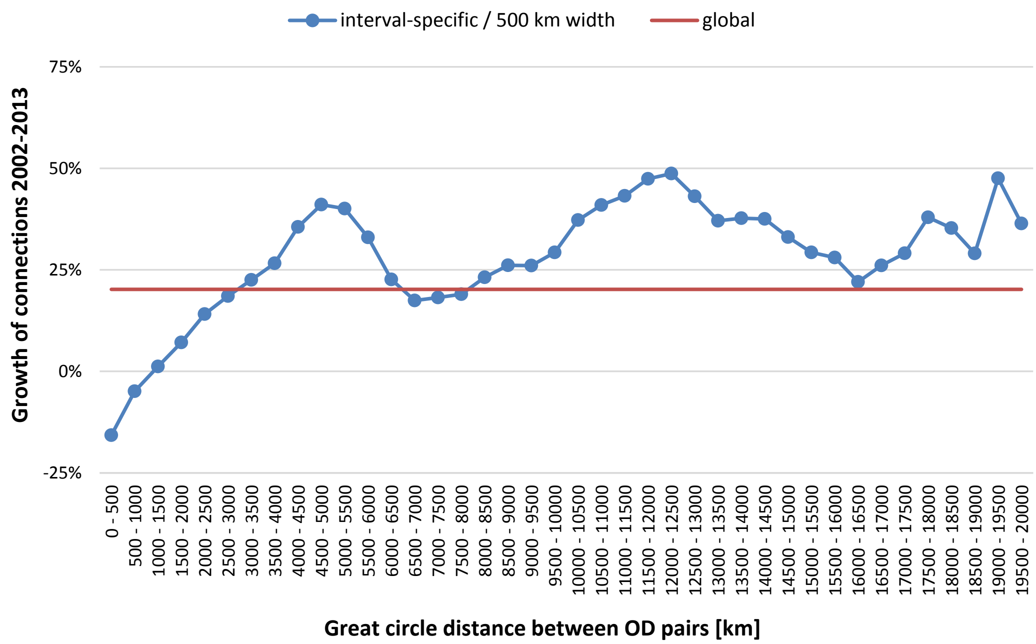

Figure 8: Global aggregate growth of number of realized demand connections compared to interval-specific growth from 2002 to 2013, interval width of 500 km

VI. Trend analysis of airfare and simplified airfare model

This section gives an overview of global average airfares from both data series (P-data and F-data). First, the average airfares are calculated for each year from reorganized ADI-data. The calculated airfares are compared to each other and the differences for the years 2009-2012 are calculated to detect systematic deviations. A correction factor of 0,984 for the preliminary data set was determined. Meaning, the P-data has systematically more expensive average airfares than the F-data. The average deviation is -1,6%. To correct the airfares of the P-Data, they are lowered by 1,6% in each year. Average airfare distributions over GCD between city pairs have been calculated and estimated with analytical functions. The parameters of these functions throughout the available time series are then correlated with the oil price as an external scenario factor [4]. Functions and correlations are used to estimate future airfares on city pairs as a function of GCD and the projected oil price at a given point of time based on the assumed scenario in a conceptual manner. The simplified global airfare model is applied to predict first order OD passenger demand networks (topology and flows) in future time steps.

ADI-data provides average one-way airfares for connections. Analogously to passenger and connections, we aim to calculate airfare distributions as a function of GCD between OD city pairs. First, the distances are divided into intervals. Airfares of all connections that belong to a certain interval are aggregated. The sum of all airfares in this interval is subsequently divided by the amount of connections, which are located in that interval. The result is the global average airfare on that specific interval:

$$average\ airfare_{interval} = \frac{\sum \mathrm{airfare}_{\mathrm{interval}}}{\#\mathrm{connections}_{\mathrm{interval}}}$$

In the following we discuss the global aggregate trend and the evolution of airfare distributions.

A. Aggregate trends

In Figure 9, the trend of the global average airfare in nominal prices as it is stated by ADI-data and the trend of inflation-adjusted, i.e. real prices are compared. Over a ten years period, from 2002-2012, the inflation-adjusted global average airfare shows an increase of 22%. In the same period though, nominal airfares increased by 55%.

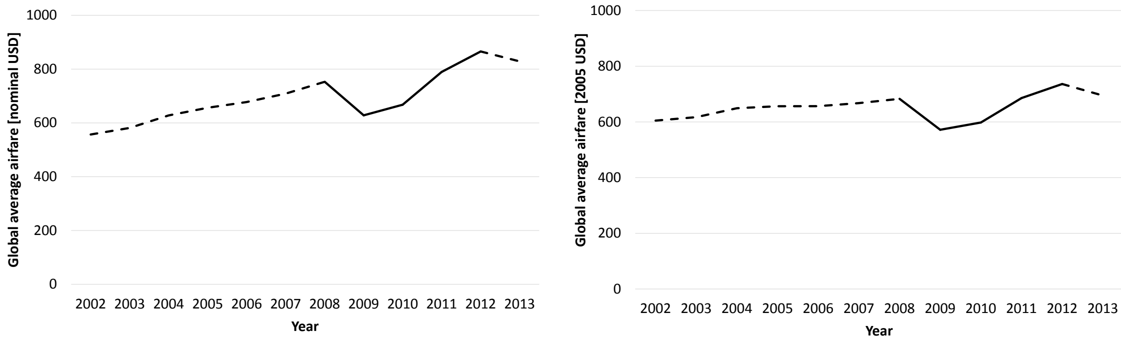

Figure 9: Trend of global average airfares, ADI nominal prices (left) and real prices in 2005 USD (right), corrected P-data and F-data mixed

B. Trend of distribution

The average airfare is calculated for every interval and plotted over the GCD (Figure 10). This procedure is applied to every year from 2002-2013. Corrected P-data and F-data are mixed to extend the observation period and get a better impression of the long-term trend. The oil price in the overall AIRCAST environment is modelled in constant 2005 USD, as is economic development in terms of GDP data. Therefore, nominal airfares as in the ADI-data, are also adjusted to constant 2005 USD using US inflation adjustment factors [9] as shown in Table 1.

Table 1: Inflation adjustment factors for conversion to constant 2005 USD [9]

Year	2002	2003	2004	**2005**	2006	2007	2008	2009	2010	2011	2012	2013
Factor	1,086	1,061	1,034	**1**	0,969	0,942	0,907	0,91	0,896	0,867	0,851	0,838

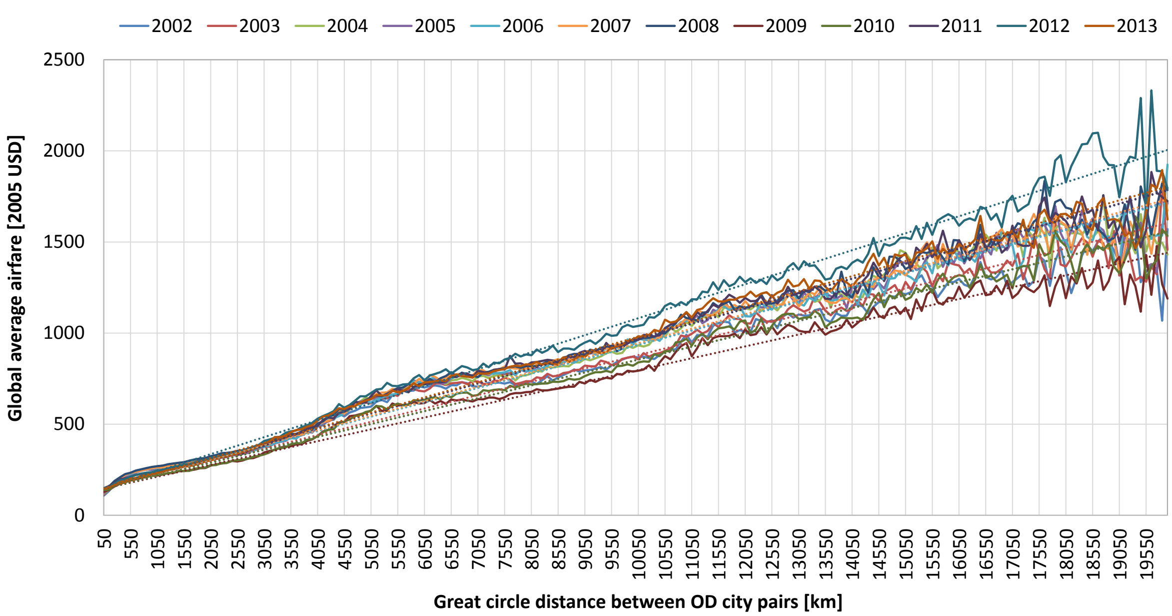

Figure 10: Time series of global average airfare distributions in constant 2005 USD with an interval width of 100 km

The average airfares increase with higher GCD between OD city pairs. It is remarkable that the correlation is almost linear. The data noise above 17500 km may be ascribed to a low number of available connections in the dataset on these intervals. From the combined P/F-data time series, we build a simplified airfare model depending on the GCD between OD city pairs and the oil price, as an external scenario factor. Inflation adjusted airfares in general, rose systematically on all distance intervals over the last 11 years with some minor volatility, not only in total aggregate of numbers, but also systematically on all intervals. This is especially visible when we compare the 2002 distribution to the 2012 distribution (see Figure 10).

The quasi-linear relationship between GCD and airfare distribution for given years is used to build a conceptual airfare model. Linear regressions are found for every year in the following form, with airfare (*AF*), the slope in a given year (*a*), the variable great circle distance (*GCD*), and a manually fixed y-intercept b_{AF}:

$$AF(year, GCD) = a\ (year) \cdot GCD + b_{AF}$$

b_{AF} is fixed 140 constant 2005 USD. The slope *a* of every year's distribution and the inflation-adjusted crude oil prices [10] are displayed in Table 2. The slope of the distribution has the unit 2005USD/km.

Table 2: Slope *a* of the airfare distributions of distinct years in 2005USD/km and Brent crude oil prices in 2005USD

Year	2002	2003	2004	**2005**	2006	2007	2008	2009	2010	2011	2012	2013
a	7,071	7,320	7,894	7,901	7,877	7,969	8,213	6,504	7,081	8,226	9,324	8,316
Oil price	27,15	30,62	39,56	54,53	63,13	68,29	87,94	56,2	71,3	96,7	94,99	91,01

Following, we plot the slopes of airfare distributions over historical oil prices as depicted in Figure 11.

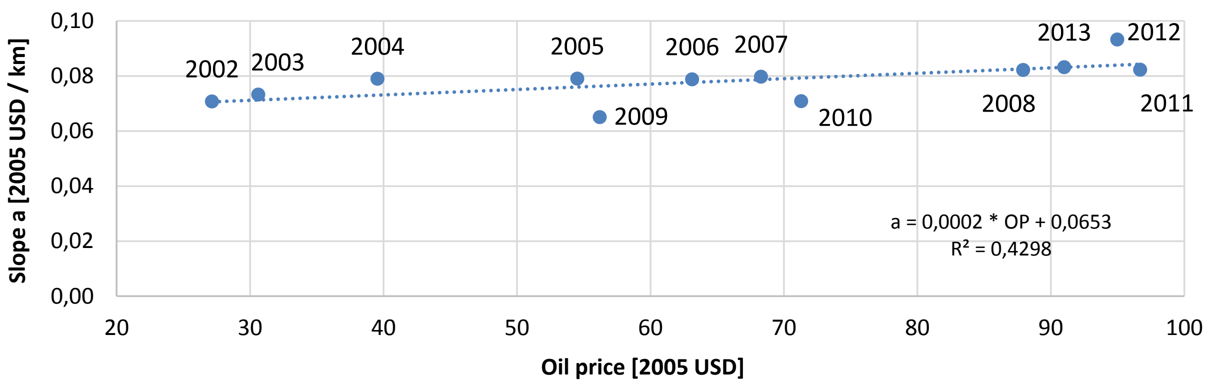

Figure 11: Correlation of the external scenario factor "oil price" and airfare slope a

We find the following functional dependence between the airfare slope and oil price *OP* in a given year:

$$a\ =\ 2 \cdot 10^{-4}\ \cdot OP + 0{,}0653$$

From the two formulas, we derive a simplified model to calculate airfare *AF* in constant 2005 USD as a function of *GCD* between OD city pairs and the oil price:

$$AF(OP, GCD) = \left(2 \cdot 10^{-4} \frac{1}{km} \cdot OP +\ 0{,}0653 \frac{USD}{km}\right) \cdot GCD + 140\ USD$$

Since the oil price changes over time in our scenarios the airfare function becomes as well time dependent. With the simplified airfare model, it is possible to estimate a future airfare on a demand city pair based on the distance between the two cities and an assumed oil price development depending on the given external scenario. This is a necessary input in the methodology to forecast the first iteration demand network.

VII. Trend analysis of network metrics

A. Directness of the global ATS

A new aggregate metric to describe the demand network and therefore passenger air traffic demand growth is introduced: Demand Passenger Kilometers (DPK). DPK in contrast to Revenue Passenger Kilometers (RPK) describe the idealized required passenger traffic performance between origin and destination over the great circle distance. The DPK is the product of the great circle distance between origin and destination of air travel and the transported volume of passengers on that distance. It is a theoretical value that describes the traffic performance if every passenger would fly on a great circle non-stop from origin to intended destination without considering actual routes flown in hub-and-spoke systems. In other words, it quantifies the ideal traffic performance of the demand network (DPK) compared to real flight segments on which aircraft are actually operated (RPK). With this metric it is possible to assess the overall directness or indirectness of the ATS by contrasting ideal DPKs with RPKs. The directness factor (DF) describes the influence of hubs and real airline routes in comparison with an ideal demand network. The DF is defined as follows:

$$DF = \frac{DPK}{RPK}$$

It can easily be seen that the reciprocal of the directness factor is equal to the average detour factor of passengers indicating the additional average kilometers a passenger is travelling due to the inefficiency of the system. We further calculate the trend of this indirectness factor over the period 2003-2013 in order to evaluate if the ATS is getting more direct or not. DPK is an aggregate metric that is necessary to describe the demand network and to derive future passenger air traffic from socio-economic boundary conditions. A change of DPKs could also be analyzed for single cities, countries or regions. When assessing the importance of cities in the global demand network, DPKs are a useful addendum compared to mere passenger volumes. DPKs consider the volume of passengers over certain distances from/to the location of interest and therefore give information about demand performance.

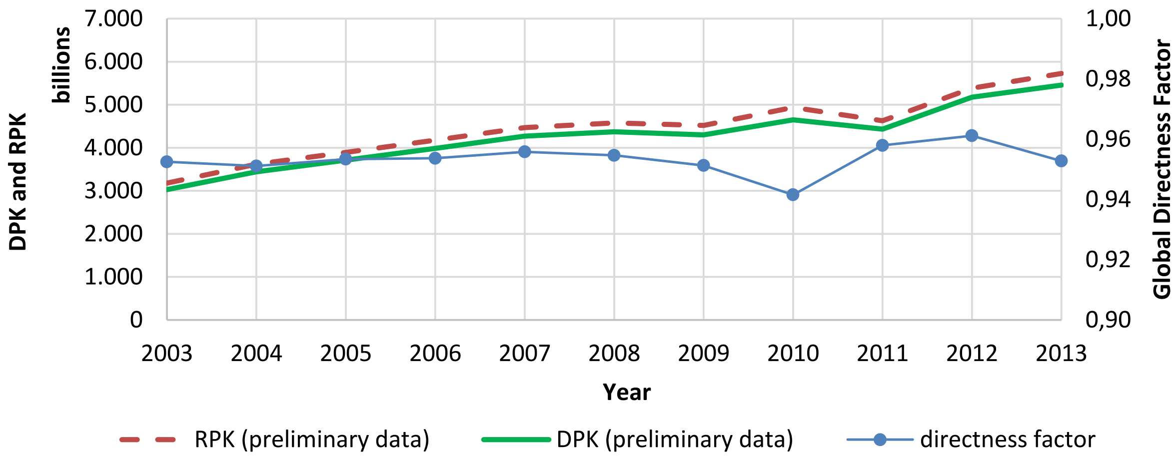

Figure 12: Trend of global demand passenger kilometer (DPK), revenue passenger kilometer (RPK) and the global directness factor (DF)

The average directness factor was calculated using ADI preliminary data only for global DPK and RPK to assure consistency. The average directness factor over the last 10 years was approximately 0,953. The overall directness of the global remained almost constant, i.e. there is no general trend towards more directness. One interpretation may be that even if there are new direct routes, old and new indirect routes are traffic-wise growing accordingly so that the overall directness of the global ATS remained more or less constant over the last 10 years.

B. Analyses of global network metrics using Gephi

The following global demand network metrics have been calculated using the software Gephi [11]. We chose the metrics that have a relevance to the quantitative understanding of a passenger demand network. Many network metrics are especially relevant for routes networks. In the following, we analyze global metrics of selected years to build time series and identify major trends. For the networks of 2002, 2006, 2009, and 2013 the global metrics Average Degree, Average Weighted Degree, Average Path Length and Network Diameter, Global Clustering Coefficient[3], Density und Modularity are evaluated. At first, we analyzed the global trend of the amount of cities in the ADI-data and the number of demand connections between cities, as depicted in Figure 13. In the past decade, the number of cities increased by 10 per year while the number of demand connections rose by about 8380 per year.

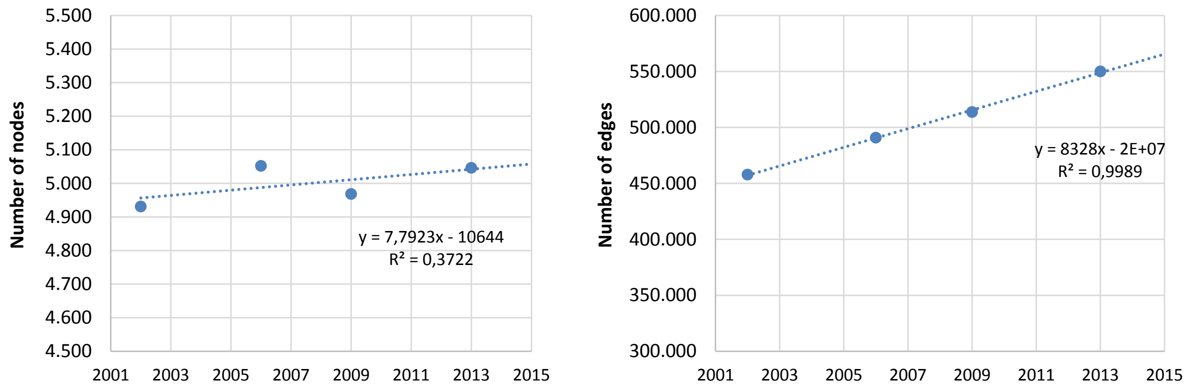

Figure 13: Trend of total number of nodes (cities) and edges (demand connections) of the global demand network

The Average Degree (see Figure 14) of the overall network is the average of all degrees in the network. A degree for a single node in a network is the number of connections that the node has within the network. [11] The Average degree is equivalent to the average amount of demand connections of a city in the global demand network. There is a rising trend of about three additional demand connections per year and city. This means that, on global average, the passengers of a city demanded three new travel destinations every year additionally to the existing destinations. The Average Weighted Degree (see Figure 14) is the global average of the Weighted Degrees of all nodes. For a single node, the Weighted Degree is the sum of weights of the edges incident to the node. In a passenger demand network, the edges are weighted with a passenger volume. Naturally, the Weighted Degree value of a city corresponds to the sum of all passengers traveling from/to that city (undirected network). The Average Weighted Degree shows a strong upward trend. In average, the amount of passengers from/to a city increases by about 45000 per year.

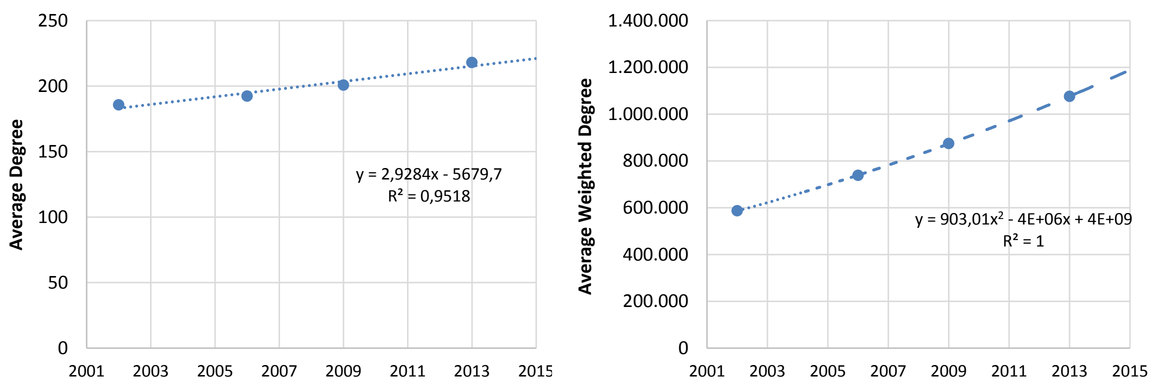

Figure 14: Trend of Average Degree and Average Weighted Degree of the global demand network

[3] in Gephi also called Average Clustering Coefficient

The Average Path Length (see Figure 15) is the average distance in terms of number of edges of all node pairs in a network. Two nodes with a direct link have a path length of one. The distance of two nodes, which have no direct link corresponds to the shortest path between the nodes. The longest path between two nodes in a network is called Network Diameter. The Network Diameter for the global passenger demand network was between 7 and 8 in the analyzed time slices. The Average Path Length concerning the demand network could be relevant for the speed of aviation-catalyzed spreading of pandemics. A lower Average Path Length implicates a faster spreading of a virus in the network. We note a slight downward trend. A virus might take in average 2 to 3 steps over intermediate cities to get from one city available in the ADI database to any other city.

The metric Density (see Figure 15) is an indicator for the completeness of a network. If all nodes in a network were connected directly, the Density would be one. [11] Concerning the passenger demand network, the Density is a theoretical value describing the actual demand connections between city pairs in comparison to all possible demand connections between all cities. We see a continuous and systematic fall of the Average Path Length and increase of the Density. The reason is that the number of overall edges (20% in 11 years) is growing much faster than the number of nodes (2% in 11 years).

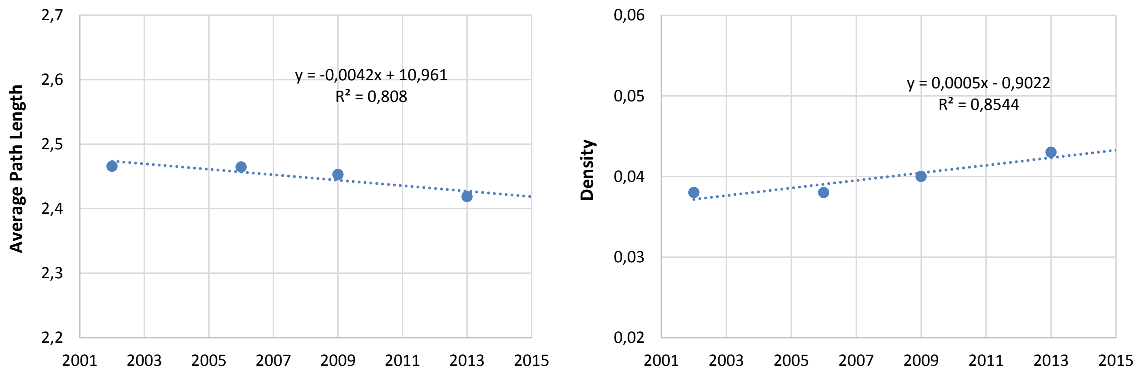

Figure 15: Trend of Average Path Length and Density of the global demand network

The Global Clustering Coefficient (see Figure 16) is a measure of the presence of triplets of nodes in a network, in which all nodes are connected in the form of a triangle. [12] In the passenger demand network, the Global Clustering Coefficient represents the average probability of the existence of a demand connection between two randomly chosen cities of the network, if they are both connected to the same third city in the network. We identified a relatively high degree of clustering in the demand network with a slightly increasing tendency over time.

The Modularity is an indicator of how much a network is divided into sub-networks. A network with a high Modularity is characterized by nodes being strongly connected within a module but nodes between different modules being weakly connected. [11] In a passenger demand network, modules correspond to communities of cities having a high number of intra-communal demand connections, but only having few or no demand connections to other communities of cities. The Modularity increased slightly in the past decade, but no significant meaning for modelling the future demand was found.

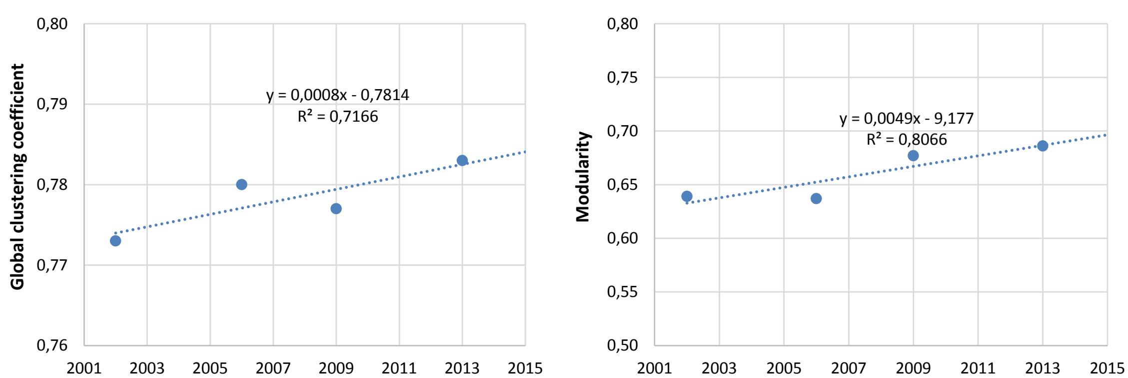

Figure 16: Trend of Global Clustering Coefficient and Modularity of the global demand network

VIII. Network visualization

As an example, global passenger demand network visualizations are displayed in Figure 17. It shows the global visualization of the demand data of the years 2002 compared to 2013. For the visualization of connections it has proven that a depiction only of demand connections with more than 100000 passengers per year is useful. The nodes represent the cities available in the ADI-data and are weighted according to the number of passengers to/from a city. The different world regions, as categorized in the ADI database are colored differently.

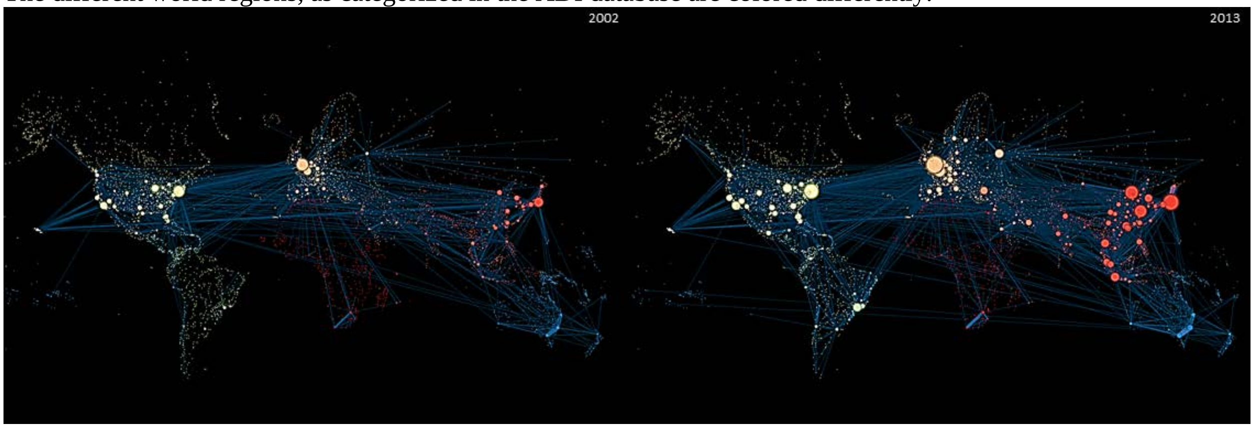

Figure 17: Sample visualizations of global demand data 2002 (left) and 2013 (right)

IX. Conclusion & outlook

We analyzed global demand data after the rearrangement to city level as backbone for future demand network forecasting, within the future air passenger traffic modelling environment AIRCAST. The identification and understanding of the development of the key metrics of the historical demand networks from 2002 to 2013 will serve as a reference for the results of generically calculated demand networks based on socio-economic scenario factors in future time slices. The gained knowledge about the trends of network metrics will be used for aggregate plausibility checks of complex model outputs. Especially valuable for this are the directness factor, average degree, connection distributions, and passenger distributions.

References

[1] Ghosh, R., Gollnick, V. “The 2° target revisited: Climate-ecological assessment loop and conceptual foundation for air transportation system re-engineering”. *Deutscher Luft- und Raumfahrt Kongress (DLRK)*, Stuttgart, Germany, 2013.

[2] Koch et al. “Climate impact assessment of varying cruise flight altitudes applying the CATS simulation approach”. *The International Conference of the European Aerospace Societies (CAES)*, 2011.

[3] Jorgen Randers. “2052: A Global Market Forecast for the Next Forty Years”. Chelsea Green Publishing, 2012.

[4] Ghosh, R., Schilling, T., Wicke, K. “Theoretical Framework of Systems Design for the Air Transportation System including an inherently quantitative Philosophy of Scenario Development”, *29th Congress of the International Council of the Aeronautical Sciences (ICAS)*, 7.-12, St. Petersburg, Russia, 2014.

[5] Terekhov, I., Ghosh, R., Gollnick, V. “A concept of forecasting origin-destination air passenger demand between global city pairs using future socio-economic development scenarios”, *53rd AIAA Aerospace Sciences Meeting*, Kissimmee, Florida, USA, 2015.

[6] Kölker, K., Bießlich, P., Lütjens, K. “FoAM - From Passenger Growth to Aircraft Movements”. *18th World Conference of the Air-Transport-Research-Society (ATRS)*, Bordeaux, France, 2014.

[7] Wenzel, S., Kölker, K., Bießlich, P., Lütjens, K. “An Approach to Forecast Air Traffic Movements at Capacity-Constrained Airports”, *14th AIAA Aviation Technology Integration and Operations Conference*, Atlanta, Georgia, USA, 2014.

[8] The World Bank. Data, Air transport, passengers carried. http://data.world-bank.org/indicator/IS.AIR.PSGR, latest access: 19 August 2014.

[9] Bureau of Labor Statistics. Databases, Tables & Calculators by Subject, http://www.bls.gov/data/inflation_calculator.htm, latest access: 20 August 2014.

[10] U.S. Energy Information Administration. Petroleum & Other Liquids. Spot Prices, http://www.eia.gov/dnav/pet/pet_pri_spt_s1_a.htm (latest access: 20 August 2014)

[11] Gephi Version 0.8.2: Statistics, program help information.

[12] Newman, M. E. J. “The structure and function of complex networks”, *Society for Industrial and Applied Mathematics*, pp. 11, 2003.